

基于逆深度自适应加权的多视图三角化方法

方维^{1*}, 杨奎²

¹北京邮电大学自动化学院, 北京 100876;

²北京航空航天大学仪器科学与光电工程学院, 北京 100191

摘要 在已知图像观测值和相机内外参数的多视图三角化中, 由于观测噪声的存在, 导致中点法和 L_2 反投影标准法分别在三角化精度和效率上存在不足。因此, 提出了一种基于逆深度自适应加权的多视图三角化方法。首先, 通过构建待估计空间三维点或多视图环境下的逆深度模型, 赋予不同视点下观测误差对应的自适应权重。然后, 确定多视图三角化近似角度误差的无偏估计模型。最后, 利用固定点迭代快速求解代价函数。在仿真和实际数据集上的实验结果表明, 本方法能很好地平衡多视图三角化的精度和效率, 且在不同噪声情况下的重建精度和迭代次数有较强的鲁棒性。

关键词 图像处理; 多视图三角化; 逆深度加权; 三维重建; 迭代优化

中图分类号 O436

文献标志码 A

doi: 10.3788/CJL202047.1204007

Inverse Depth Adaptive Weighting Based Multi-View Triangulation Method

Fang Wei^{1*}, Yang Kui²

¹School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China;

²School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

Abstract In the multi-view triangulation with known image observation values and camera internal and external parameters, due to the existence of observation noise, the midpoint method and the L_2 back projection standard method have insufficient triangulation accuracy and efficiency, respectively. Therefore, this paper proposes an inverse depth adaptive weighting based multi-view triangulation method. First, by constructing an inverse depth model of the three-dimensional points to be estimated in a multi-view environment, the corresponding weights to the observation errors are assigned under different viewpoints. Then an unbiased estimation model of the approximate angle error for the multi-view triangulation is determined. Finally, a fixed-point iteration is carried out to quickly solve the cost function. Experimental results both on simulation and real datasets show that the proposed method can obtain a better balance between accuracy and efficiency for multi-view triangulation, and the reconstruction accuracy and the number of iterations under different noise conditions are robust.

Key words image processing; multi-view triangulation; inverse depth weighting; three-dimensional reconstruction; iterative optimization

OCIS codes 100.2000; 100.5010; 150.6910

1 引言

从多幅图像中恢复得到场景的三维(3D)结构, 是机器人^[1]、增强/虚拟现实^[2]和视觉测量^[3-5]等应用领域中的关键技术。已有方法^[6-8]虽然能通过大量二维(2D)图像, 恢复得到高精度的三维场景, 但

如何快速准确地从不同视点下的图像中恢复得到三维场景结构, 依然是三维视觉领域值得研究的问题。因此, 需要在假设相机参数和特征匹配准确的前提下, 研究多视图的三角化方法。

典型的多视图三角化方法通过不同的相机内参数和位姿构建对应的投影矩阵, 然后利用线性化方

收稿日期: 2020-06-18; 修回日期: 2020-07-20; 录用日期: 2020-07-28

基金项目: 国家重点研发计划(2019YFC0119200)、北京市自然科学基金(3204050)、虚拟现实技术与系统国家重点实验室开放基金(VRLAB2020B05)、中央高校基本科研业务费(2019RC26)

* E-mail: fangwei@bupt.edu.cn

法求解 $\mathbf{AX} = \mathbf{0}$ ^[9], 其中, 矩阵 \mathbf{A} 由多视图中匹配的特征点和投影矩阵组成, 然后以非迭代的代数方法求解三角化结果 \mathbf{X} , 但该方法易受图像观测噪声的影响^[10], 数值稳定性较低。

为提高三角化的精度, Hartley 等^[11]根据图像特征点和极线几何间的约束, 提出了基于最优解的二视图三角化方法。在三视图的最优解三角化中, Stewenius 等^[12]用计算交互代数的方法解算多项式矩阵, 通过求解 47×47 矩阵的特征向量获取三角化结果。但上述最优解三角化方法中, 待求解多项式的自由度随图像数量的增加呈指数增长, 导致计算复杂度急剧增加。且现有最优解方法主要应用在二视图^[11]和三视图^[13]的三角化中, 对于大于三幅图像的多视图场景, 尚无通用的最优解三角化方法。

为实现 $N > 3$ 的多视图三角化, 人们提出了基于 L_1 的三角化方法, 但该方法对相机成像模型^[14]和图像噪声^[15]较为敏感, 难以获得全局最优的三角化结果。Agarwal 等^[16]用分数规划方法解决了多视图几何中的全局三角化问题, 在假设图像存在高斯噪声的情况下, 利用分支界定算法找到逼近三角化的全局最优解。Dai 等^[17]提出了一种基于 L_∞ 的优化方法, 通过逐步缩小凸区域的方式求取三角化结果, 提高了 L_∞ 方法对多视图的三角化效率。Zhang 等^[18]通过在 L_∞ 多视图三角化中, 选取部分有代表性的观测向量作为待求解任务的子集, 然后对子集数据进行求解, 以逼近全局多视图的三角化结果, 有效提高了 L_∞ 方法在大规模多视图三角化中的效率。但该方法只选用了部分多视图图像子集, 没有充分利用所有图像的观测信息, 且对观测噪声极为敏感^[19]。

在实际多视图三角化应用中, 基于 L_2 优化方法虽不能保证获取全局最优的三维点, 但当观测误差服从正态分布时, 其优化准则服从最大似然估计, 能在获取可信初始估计的基础上, 通过持续迭代逼近最优解, 平衡三角化的效率和精度^[20], 广泛应用于运动恢复结构 (SfM)^[7]与同步定位和建图 (SLAM)^[21-22]系统中。为提高 L_2 三角化的效率, Yang 等^[23]在中点法的基础上, 构建了点到相机距离的加权优化模型, 在保证图像反投影误差精度的基础上提高了多视图三角化的效率。但该方法没有进一步讨论噪声对三角化结果的影响, 且缺乏多视图三角化所得场景点空间精度的分析。此外, 当前 L_2 反投影误差的三角化方法, 主要应用于透视相机模型, 对于鱼眼和全景相机模型 (相机畸变系数较

大), 在图像边缘附近的特征点, 基于传统图像反投影误差方式难以准确逼近最优的三角化结果, 导致该方法对不同相机类型的适应性较差^[24]。因此, 如何快速高效地实现环境适应性高的多视图三角化方法, 还需进一步研究。

针对上述问题, 本文提出了一种基于逆深度自适应加权的多视图三角化方法, 通过构建空间三维点在不同视角下的逆深度权重模型, 研究了具有深度感知能力的多视图三角化优化模型。利用固定点迭代方法对空间三维点进行快速求解, 有效平衡了多视图三角化的精度和效率, 并在几何上近似于基于角度误差的三角化方法, 在不同相机模型下具有更好的适用性。

2 多视图三角化问题描述

对于待三角化的空间点 $\mathbf{X} \in \mathbf{R}^3$, 假设其能被不同位置下的相机 \mathbf{C}_i 可见, 根据该三维点 \mathbf{X} 在第 i 幅图像平面上的观测值 \mathbf{x}_i , 可获取从相机中心指向图像观测点的单位观测向量 \mathbf{w}_i , 如图 1 所示, 其中, θ_i 为 \mathbf{w}_i 与向量 $\mathbf{X} - \mathbf{C}_i$ 间的夹角。不同视图下的观测表达式为

$$\mathbf{X} = \mathbf{C}_i + d_i \mathbf{w}_i, \quad (1)$$

式中, d_i 为空间点 \mathbf{X} 到相机光心 \mathbf{C}_i 的深度值。理论情况下, 多视图三角化中不同的观测向量 \mathbf{w}_i 会在三维空间中交于点 \mathbf{X} 。

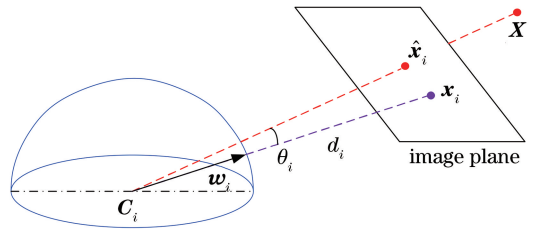


图 1 三角化描述示意图

Fig. 1 Schematic diagram of triangulation description

由于观测噪声的存在, 以 \mathbf{w}_i 为方向、 d_i 为长度的观测向量, 无法准确相交于同一空间点。因此, 在实际的 N 视图三角化中, 需要通过计算空间点 \mathbf{X} 在像平面上的投影点 $\hat{\mathbf{x}}_i$, 求解反投影误差平方和 $\sum_{i=1}^N \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2$ 的最小值, 以逼近空间三维点 \mathbf{X} 。

为简化计算, 用观测向量 \mathbf{w}_i 和相机中心到空间点 \mathbf{X} 方向向量间的夹角误差 $\angle(\mathbf{w}_i, \mathbf{X} - \mathbf{C}_i)$ 表征图像误差的优化模型。对于 N 视图的三维重建场景, 得到的待优化三角化求解模型为

$$C(\mathbf{X}) = \sum_{i=1}^N [\angle(\mathbf{w}_i, \mathbf{X} - \mathbf{C}_i)]^2. \quad (2)$$

根据空间点 \mathbf{X} 在观测向量 \mathbf{w}_i 上的正交投影矩阵 $\mathbf{w}_i \mathbf{w}_i^T$, 计算空间点 \mathbf{X} 到观测向量 \mathbf{w}_i 的距离 $f_i(\mathbf{X}) = \|(\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T)(\mathbf{X} - \mathbf{C}_i)\|$, $\|\cdot\|$ 为矩阵的 1 范数, 进而得到基于多视图三角化的线性求解模型

$$C(\mathbf{X}) = \sum_{i=1}^N f_i^2(\mathbf{X}) = \sum_{i=1}^N \|(\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T)(\mathbf{X} - \mathbf{C}_i)\|^2. \quad (3)$$

(3)式作为多视图三角化的代价函数, 可通过矩阵的乘法和对 3×3 矩阵的逆运算, 高效完成多视图三角化的求解^[25]。

3 逆深度自适应加权三角化

3.1 逆深度自适应加权误差模型

上述线性多视图三角化求解方法, 虽然重建效

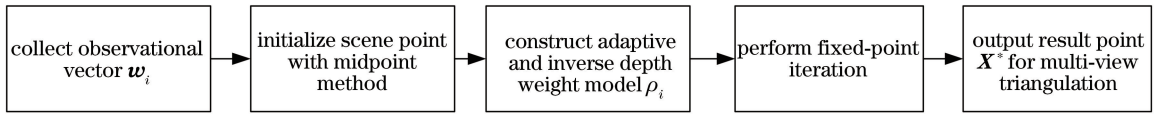


图 2 本方法的实现流程

Fig. 2 Implementation process of our method

在基于中点法所得三角化结果基础上, 构建具有逆深度自适应特性的加权模型, 可表示为

$$\rho_i = 1/d_i. \quad (4)$$

对于第 i 个相机, 其三维点的深度值可表示为 $d_i = \|\mathbf{X} - \mathbf{C}_i\|$, 该权重参数可在后续三角化过程中, 根据空间三维点 \mathbf{X} 的迭代变化进行最优估计的动态逼近。结合(3)式和(4)式, 得到第 i 个视图下具有距离自适应特性的求解函数 $g_i(\mathbf{X})$

$$g_i(\mathbf{X}) = \rho_i f_i(\mathbf{X}). \quad (5)$$

对于 N 视图的三角化问题, 构建逆深度加权后的代价函数可表示为

$$\hat{C}(\mathbf{X}) = \sum_{i=1}^N g_i^2(\mathbf{X}) = \sum_{i=1}^N \|\rho_i (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T)(\mathbf{X} - \mathbf{C}_i)\|^2. \quad (6)$$

在逆深度自适应加权后的误差模型 $g_i(\mathbf{X})$ 中, 结合图 1 的几何投影关系可知, 逆深度加权后的待求解函数 $g_i(\mathbf{X})$ 可表示为 θ_i 的正弦值。当三角化误差较小时, 可认为 $\sin \theta_i \approx \theta_i$, 即加权后的代价函数可近似逼近角度误差的平方和, 避免了(3)式中直线 $\mathbf{C}_i \mathbf{X}$ 上各点在图像平面上投影导致的有偏现象。

上述逆深度加权后的代价函数 $\hat{C}(\mathbf{X})$ 可有效逼

率较高, 且在不同相机与空间三维点的距离彼此相近时具有较好的三角化精度^[26]; 但在通用场景下的多视图三角化中, 相机空间位置分布的随机性和观测噪声的存在, 导致该线性三角化方法所得结果误差较大。如图 1 所示, 当空间点 \mathbf{X} 沿直线 $\mathbf{C}_i \mathbf{X}$ 朝远离相机中心 \mathbf{C}_i 方向运动时, 该点到观测向量的距离误差也会随之增加。但对于第 i 个相机而言, 所有直线 $\mathbf{C}_i \mathbf{X}$ 上的三维点在图像平面上的投影点始终为 $\hat{\mathbf{x}}_i$, 导致(3)式中的三角化求解方法是有偏的, 尤其是当该三维点的深度值较小时, 会出现较大的反投影误差。同时, 当不同相机到待测空间点的距离变化较大时, 其三角化结果也会出现较大的误差。

为补偿各视图在不同景深变化时对三角化精度的影响, 提出了一种基于逆深度自适应加权的多视图三角化方法, 具体流程如图 2 所示。

近基于角度误差的三角化方法, 考虑到角度误差具有的旋转不变性^[27], 本方法除了可在透视相机的多视图三角化中应用外, 也可适用于广角或鱼眼等相机模型。由于赋予了逆深度加权参数, (6)式不再是一个线性最小二乘优化的问题, 增加了后续的求解复杂性。

3.2 逆深度自适应加权的多视图三角化求解

为从逆深度加权后的代价函数 $\hat{C}(\mathbf{X})$ 中恢复得到三角化的空间点 \mathbf{X} , 首先求解该代价函数的梯度矩阵 $\nabla \hat{C}(\mathbf{X})$, 可表示为

$$\nabla \hat{C}(\mathbf{X}) = \frac{\partial \hat{C}(\mathbf{X})}{\partial \mathbf{X}} = 2 \sum_{i=1}^N \{ \rho_i^2 (\mathbf{X} - \mathbf{C}_i)^T \cdot [(\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T) - \hat{C}(\mathbf{X}) \mathbf{I}] \}, \quad (7)$$

令 $\nabla \hat{C}(\mathbf{X}) = 0$, 得到

$$\sum_{i=1}^N [\rho_i^2 (\mathbf{X} - \mathbf{C}_i)^T (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T)] = \sum_{i=1}^N [\rho_i^2 (\mathbf{X} - \mathbf{C}_i)^T \hat{C}(\mathbf{X})]. \quad (8)$$

由(8)式可知, 对于待估计的空间最优点 \mathbf{X}^* , 满足 $\nabla \hat{C}(\mathbf{X}^*) = 0$, 进一步整理得到固定点迭代函

数为

$$\left\{ \sum_{i=1}^N [\rho_i^2 (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T)] \right\} \mathbf{X}^* = \sum_{i=1}^N [\rho_i^2 \hat{\mathbf{C}}(\mathbf{X}^*) \cdot (\mathbf{X}^* - \mathbf{C}_i) + \rho_i^2 (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T) \mathbf{C}_i]. \quad (9)$$

假设多视图三角化的初值 \mathbf{X}_0 已知, 根据相机中心坐标 \mathbf{C}_i 和对应的测量向量 \mathbf{w}_i , 可得到

$$\mathbf{X}_n = \mathbf{M}^{-1} \sum_{i=1}^N [\rho_i^2 \hat{\mathbf{C}}(\mathbf{X}_{n-1}) \cdot (\mathbf{X}_{n-1} - \mathbf{C}_i) + \rho_i^2 (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T) \mathbf{C}_i], \quad (10)$$

式中, $\mathbf{M} = \sum_{i=1}^N [\rho_i^2 (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T)]$. 首先, 通过(3)式的线性中点法解得空间点 \mathbf{X} 的初始估计 \mathbf{X}_0 . 然后, 根据(10)式进一步得到第 n 次迭代更新后的三角化结果 \mathbf{X}_n . 直到 $\|\mathbf{X}_n - \mathbf{X}_{n-1}\|$ 小于某一阈值 T 时 (实验中设 $T = 0.001$), 得到多视图三角化的最优解 \mathbf{X}^*

$$\mathbf{X}^* = \mathbf{X}_n. \quad (11)$$

矩阵 \mathbf{M} 的特征值为 $\{\rho_i^2, \rho_i^2, 0\}$, 特征值为 0 对应的特征向量为观测向量 \mathbf{w}_i , 且迭代过程中的逆深度值 ρ_i 始终大于 0. 因此, 只需所有观测向量 $\{\mathbf{w}_i\}_1^N$ 不完全平行, 矩阵 \mathbf{M} 的各分量特征值都为正, 进而保证模型可以稳定迭代收敛至三角化的最优解.

4 实验验证

采用仿真和实际公开数据集进行多视图三角化实验, 选取经典的中点法 (Midpoint) 和当前主流应用的 L_2 反投影误差标准法 (L_2 Rep) 作为对比基准, 以验证本方法的有效性. 虽然中点法的三角化精度较低, 但其效率是目前三角化方法中最高的^[9]. L_2

反投影误差法通过构建基于图像反投影误差的优化模型, 利用梯度下降迭代求解三角化结果. 该方法能有效平衡三角化的精度与效率, 广泛应用于当前主流的 SfM^[7] 和 SLAM^[21] 系统中. 所有方法均通过中点法得到多视图三角化的初值, 实验使用的软件为 Matlab, CPU 为 Intel i7-8550U, 计算机的内存为 8 G.

4.1 仿真数据验证

在仿真数据集中, 图像的尺寸为 1024 pixel \times 1024 pixel, 焦距为 400 pixel, 仿真相机的内参数矩阵为

$$\mathbf{K} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 400 & 0 & 512 \\ 0 & 400 & 512 \\ 0 & 0 & 1 \end{bmatrix}, \quad (12)$$

式中, (f_u, f_v) 为相机归一化焦距, (u_0, v_0) 为相机的主点坐标. 根据多视图三角化的典型应用场景, 生成的三种仿真数据集如图 3 所示. 其中, 每组仿真数据集中相机的数量为 100, 三维点的数量为 3000. 具体仿真场景描述如下.

1) 相机和三维点在空间随机分布, 该仿真场景能模拟根据不同图像数据源执行多视图三角化的过程 (type A), 如图 3(a) 所示.

2) 相机沿着直线轨迹运动并观测三维场景, 该仿真场景能模拟摄影测量中通过航拍图像恢复现场三维结构的过程 (type B), 如图 3(b) 所示.

3) 相机沿着圆形轨迹围绕三维空间点进行运动, 该仿真场景能模拟针对特定目标开展基于多视图三角化的过程 (type C), 如图 3(c) 所示.

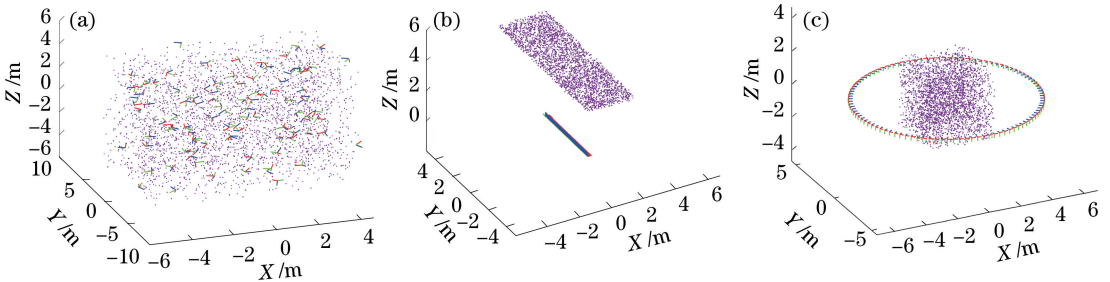


图 3 多视图三角化仿真数据集. (a) Type A; (b) type B; (c) type C

Fig. 3 Synthetic datasets for multi-view triangulation. (a) Type A; (b) type B; (c) type C

仿真时, 首先, 在相机参数已知的情况下, 根据模拟的相机位姿和空间三维点坐标, 将场景中的三维点向相机平面进行投影, 得到其在不同像平面上的二维投影点坐标, 并赋予投影点相应的高斯噪声, 以此作为空间三维点在不同相机位姿下的图像观测

值. 其次, 根据已知的相机中心和观测向量, 利用不同的多视图三角化方法重建该空间三维点. 最后, 统计不同多视图三角化方法的执行时间, 重建三维点在图像平面上的 2D 反投影误差及其与理论空间三维点间的欧氏距离, 评估不同多视图三角化方法

的精度和效率。不同多视图三角化方法的实验结果如表 1 所示,考虑到仿真实验的一般性,表中的数据是通过 30 次重复实验结果的平均值。为进一步评

估本方法在不同噪声环境下的鲁棒性,分别在理论反投影图像点上叠加 σ 为 6, 12, 24 pixel 的高斯噪声。

表 1 基于仿真数据的多视图三角化方法得到的结果

Table 1 Results obtained by the multi-view triangulation method based on simulation data

σ / pixel	Method	Type A			Type B			Type C		
		Time / s	3D error / m	2D error / pixel	Time / s	3D error / m	2D error / pixel	Time / s	3D error / m	2D error / pixel
6	Midpoint	0.648	0.023	6.107	2.521	0.021	5.945	0.853	0.017	5.825
	Ours	1.761	0.017	5.728	5.971	0.018	5.938	2.251	0.016	5.801
	L_2 Rep	2.996	0.015	5.704	11.668	0.017	5.936	4.005	0.016	5.797
12	Midpoint	0.617	0.045	12.244	2.462	0.059	11.941	0.810	0.036	11.652
	Ours	1.802	0.031	11.669	5.675	0.037	11.883	2.068	0.031	11.595
	L_2 Rep	2.841	0.030	11.620	11.678	0.035	11.879	3.933	0.029	11.582
24	Midpoint	0.603	0.094	25.953	2.578	0.208	24.278	0.849	0.078	23.498
	Ours	1.914	0.063	23.009	7.191	0.074	23.781	2.382	0.064	23.216
	L_2 Rep	3.224	0.059	22.905	14.965	0.071	23.773	4.017	0.059	23.180

从表 1 可以发现,虽然中点法在多视图三角化中具有明显的速度优势,但其三角化的精度最低。与中点法相比,本方法能获得更高的重建精度;与 L_2 Rep 反投影误差标准方法相比,本方法在保证重建精度的前提下,将重建效率约提升了 1 倍,能更好地平衡多视图三角化过程中的效率和精度。此外,随着图像观测噪声的增加,三种方法的三角化时间基本保持不变,但本方法在不同噪声情况下均能得到较好的空间三维点精度。

环境下具有快速稳定收敛的能力。

4.2 实际数据验证

在中点法所得初值的基础上,为实现多视图三角化精度的优化,在上述不同数据集下所需要的迭代次数如图 4 所示。可以发现,对于模拟得到的不同多视图应用场景,本方法获取最优解所需的迭代次数保持高度一致。其次,当图像观测噪声增加时,本方法所需的迭代次数也相应增加,但其增加的幅度逐渐放缓。即使叠加的高斯噪声达到 30 pixel(约为图像尺寸的 3%),本方法平均只需要迭代 3.25 次,即可收敛至最优三维点估计,进一步说明了本方法在不同噪声

为进一步验证本方法的有效性,借助于瑞典 Lund 大学提供的公开数据集^[28]进行实验。该数据集包含了不同场景下大量的图像数据,并提供了相机参数和不同视觉特征间的匹配关系。基于该数据集,用不同的多视图三角化方法对不同场景进行三维重建,结果如表 2 所示。其中, Dataset ID 1~10 分别表示 Lund Cathedral、Aos Hus、San Marco、Orebro Castle、Buddah Statue、East Indiaman Goteborg、Ystad Monestary、Round Church、Skansen Kronan、Skansen Lejonet 数据集。考虑到该数据集中缺乏空间三维点的基准坐标值,只根据 2D 反投影误差对三角化结果进行对比分析。实验结果表明,本方法获取的场景反投影误差和基于 L_2 反投影标准方法基本一致,优于中点法。在三角化效率上,本方法明显优于 L_2 反投影标准方法。对不同的公开数据集,由本方法得到的多视图三角化结果如图 5 所示。

在上述公开数据集提供图像观测的基础上,叠加 σ 为 5 pixel 的高斯噪声,以进一步验证本方法在实际场景中的鲁棒性。不同方法得到的三角化结果如表 3 所示。对比表 2 可以发现,不同方法的执行时间几乎保持不变。在三角化精度方面,所得三维点的 2D 反投影误差随外加噪声而有所增加。但在不同噪声情况下,本方法和标准的 L_2 反投影优化方法在三角化结果上基本保持一致,明显高于中点法,这表明本方法在不同实际噪声环境下的鲁棒性较强。

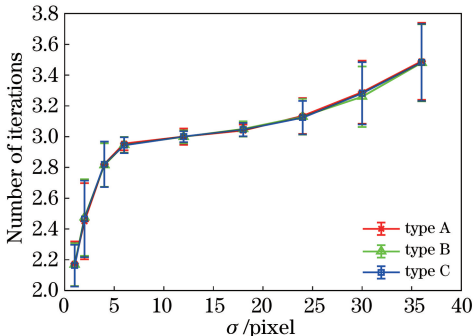


图 4 不同噪声水平下的迭代情况

Fig. 4 Iteration performance at different noise levels

表 2 不同方法基于 Lund 公开数据集的结果

Table 2 Results of different methods based on Lund's public dataset

ID	Dataset		Time /s			Mean 2D reprojection errors /pixel					
	View	Point	Midpoint	Ours	L_2 Rep	Mean			Variance		
						Midpoint	Ours	L_2 Rep	Midpoint	Ours	L_2 Rep
1	1208	159055	23.082	43.014	91.910	1.088	1.078	1.077	0.217	0.206	0.205
2	800	354134	22.303	45.273	93.063	0.816	0.805	0.805	0.302	0.289	0.287
3	1498	231507	33.786	66.944	134.205	0.807	0.799	0.798	0.331	0.316	0.315
4	761	53857	10.998	20.410	40.532	0.942	0.936	0.936	0.196	0.191	0.190
5	322	156356	9.440	15.233	31.873	0.651	0.649	0.649	0.270	0.266	0.265
6	179	25655	2.723	5.736	11.040	1.127	1.122	1.121	0.386	0.377	0.376
7	290	139951	14.520	26.328	53.460	0.970	0.968	0.967	0.170	0.169	0.168
8	92	84643	6.366	11.699	22.445	0.387	0.385	0.385	0.121	0.119	0.119
9	131	28371	3.812	7.254	14.793	0.807	0.802	0.802	0.175	0.170	0.169
10	368	74423	8.370	15.871	34.350	1.032	1.023	1.022	0.209	0.202	0.200

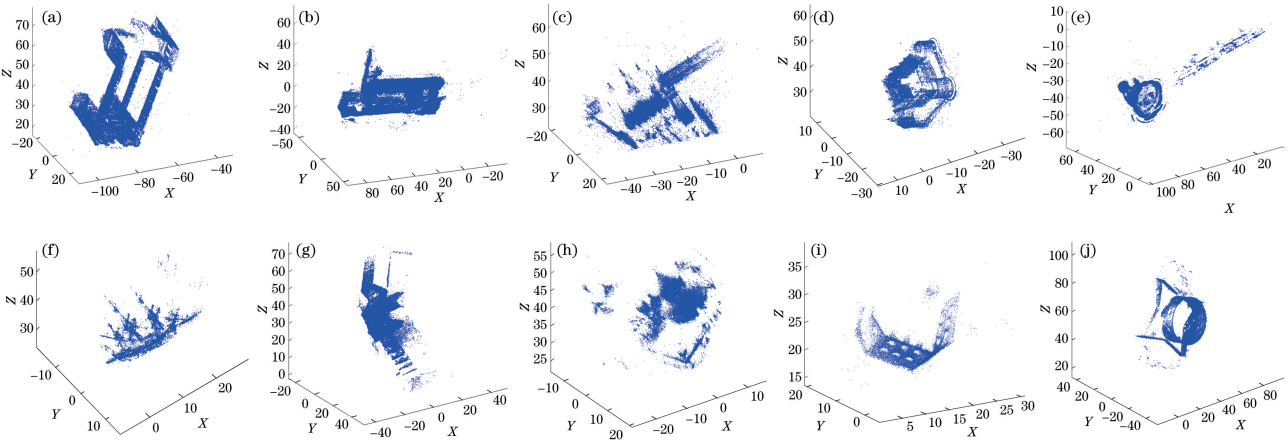


图 5 本方法在公开数据集下的多视图三角化结果。(a) Lund Cathedral; (b) Aos Hus; (c) San Marco; (d) Orebro Castle; (e) Buddah Statue; (f) East Indiaman Goteborg; (g) Ystad Monestary; (h) Round Church; (i) Skansen Kronan; (j) Skansen Lejonet

Fig. 5 Multi-view triangulation results of our method under the public datasets. (a) Lund Cathedral; (b) Aos Hus; (c) San Marco; (d) Orebro Castle; (e) Buddah Statue; (f) East Indiaman Goteborg; (g) Ystad Monestary; (h) Round Church; (i) Skansen Kronan; (j) Skansen Lejonet

表 3 不同方法在 Lund 公开数据集的实验结果($\sigma=5$ pixel)

Table 3 Experimental results of different methods on the Lund public dataset ($\sigma=5$ pixel)

ID	Time /s			Mean 2D reprojection error /pixel		
	Midpoint	Ours	L_2 Rep	Midpoint	Ours	L_2 Rep
1	20.508	44.275	90.556	2.136	2.105	2.103
2	23.829	46.673	96.919	2.964	2.934	2.930
3	34.913	63.022	139.873	2.663	2.627	2.624
4	11.380	20.585	46.909	1.677	1.664	1.662
5	9.691	17.316	38.984	3.050	3.046	3.044
6	2.724	5.838	11.809	2.356	2.350	2.351
7	13.903	27.213	58.175	2.257	2.241	2.236
8	6.289	12.397	26.147	2.829	2.813	2.811
9	3.859	6.827	15.900	1.917	1.871	1.870
10	7.786	15.550	34.184	2.139	2.109	2.105

本方法基于 Lund 公开数据集在不同噪声情况下的平均迭代次数如图 6 所示,可以发现,在无额外叠加噪声的情况下,本方法的平均迭代次数约为

2.5,且在不同场景下保持高度一致。对图像观测叠加 $\sigma=5$ pixel 的高斯噪声时,本方法的平均迭代次数增加至 2.9 次,与未叠加观测噪声的迭代次数差

别不大。这表明本方法在不同场景和不同噪声水平下,均具有稳定高效收敛的能力。

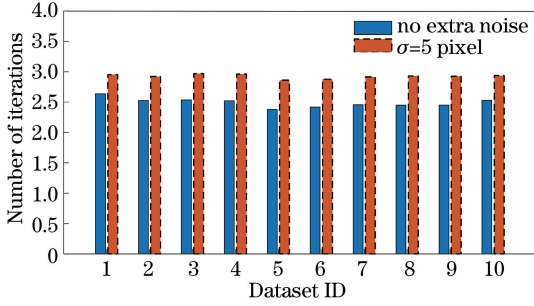


图 6 本方法在公开数据集中多视图三角化的迭代次数

Fig. 6 Number of iterations of our method for multi-view triangulation in public datasets

5 结 论

在多视图三维重建中,提出了一种基于逆深度自适应加权的多视图三角化方法,通过构建待估计三维点的逆深度权重优化模型,在求解过程中动态调整代价函数的有偏估计特性,有效克服了传统线性三角化方法重建误差较大的问题。实验结果表明,相比 L_2 反投影标准方法,本方法在保证重建精度的前提下提高了三角化效率。此外,通过构建逆深度自适应加权的优化模型,多视图重建的精度近似于基于角度误差的三角化方法,角度误差具有的旋转不变性,使本方法可适用于不同的相机模型,能有效平衡多视图三角化的精度和效率,在不同场景和噪声环境下具有快速鲁棒的迭代收敛能力。但本方法依赖线性中点法的结果作为初值,为评价本三角化方法在不同噪声水平下的鲁棒性,后续还将进一步探索不同初值误差对三角化精度优化的影响,拓展本方法在视觉定位等领域的应用。

参 考 文 献

[1] Campos R, Garcia R, Alliez P, et al. A surface reconstruction method for in-detail underwater 3D optical mapping [J]. *The International Journal of Robotics Research*, 2015, 34(1): 64-89.

[2] Han L, Zheng T, Zhu Y H, et al. Live semantic 3D perception for immersive augmented reality[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 26(5): 2012-2022.

[3] Wang G, Shang Y, Guan B L, et al. Flexible calibration of setting relation of a multi-camera rig for non-overlapping views[J]. *Chinese Journal of Lasers*, 2017, 44(6): 0604004.

王刚, 尚洋, 关棒磊, 等. 无重叠视场多相机组安装

关系的灵活标定[J]. *中国激光*, 2017, 44(6): 0604004.

[4] Qiao Y J, Tan S Z, Jiang J G. Planning strategy for multi-visual measurement networking[J]. *Acta Optica Sinica*, 2018, 38(5): 0515005.

乔玉晶, 谭世征, 姜金刚. 一种多视觉测量组网规划策略[J]. *光学学报*, 2018, 38(5): 0515005.

[5] Cui H X, Chen L J, Wang N, et al. Calibration method with implicit constraints for multi-view combined camera using automatic coding of marker points[J]. *Chinese Journal of Lasers*, 2020, 47(1): 0110003.

崔红霞, 陈丽君, 王宁, 等. 基于标志自动编码的多视组合相机隐式约束标定方法[J]. *中国激光*, 2020, 47(1): 0110003.

[6] Wu C C. Towards linear-time incremental structure from motion[C] // 2013 International Conference on 3D Vision, June 29-July 1, 2013, Seattle, WA, USA. New York: IEEE, 2013: 127-134.

[7] Schönberger J L, Frahm J M. Structure-from-motion revisited[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4104-4113.

[8] Xie Z X, Zhou Z Q. Method of space point positioning based on structure-from-motion [J]. *Laser & Optoelectronics Progress*, 2018, 55(8): 081502.

解则晓, 周作琪. 基于运动恢复结构的空点定位方法[J]. *激光与光电子学进展*, 2018, 55(8): 081502.

[9] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. 2nd ed. Cambridge: Cambridge University Press, 2004

[10] Wang Y X, Liu Y. Robust structure from motion approach based on multi cameras[J]. *Chinese Journal of Scientific Instrument*, 2015, 36(1): 230-238.

王宇希, 刘越. 一种基于多摄像机的鲁棒运动结构重建方法[J]. *仪器仪表学报*, 2015, 36(1): 230-238.

[11] Hartley R I, Sturm P. Triangulation[J]. *Computer Vision and Image Understanding*, 1997, 68(2): 146-157.

[12] Stewenius H, Schaffalitzky F, Nister D. How hard is 3-view triangulation really? [C] // 2005 IEEE International Conference on Computer Vision, October 17-21, 2005, Beijing, China. New York: IEEE, 2005: 686-693.

[13] Lindstrom P. Triangulation made easy [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 1554-1561.

- [14] Ke Q F, Kanade T. Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, CA, USA. New York: IEEE, 2005: 739-746.
- [15] Hess-Flores M, Recker S, Joy K I. Uncertainty, baseline, and noise analysis for L_1 error-based multi-view triangulation [C] // 2014 22nd International Conference on Pattern Recognition, August 24-28, 2014, Stockholm, Sweden. New York: IEEE, 2014: 4074-4079.
- [16] Agarwal S, Chandraker M K, Kahl F, et al. Practical global optimization for multiview geometry [J]. International Journal of Computer Vision, 2008, 79(3): 271-284.
- [17] Dai Z J, Wu Y H, Zhang F J, et al. A novel fast method for L_∞ problems in multiview geometry [M]//Fitzgibbon A, Lazebnik S, Perona P, et al. Computer Vision-ECCV 2012. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, 7576: 116-129.
- [18] Zhang Q G, Chin T J. Coresets for triangulation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(9): 2095-2108.
- [19] Zhou G Q, Wang Q, Xiao Z L. Robust outlier removal using penalized linear regression in multiview geometry [J]. Neurocomputing, 2017, 267: 455-465.
- [20] Hartley R, Kahl F, Olsson C, et al. Verifying global minima for L_2 minimization problems in multiple view geometry[J]. International Journal of Computer Vision, 2013, 101(2): 288-304.
- [21] Mur-Artal R, Tardós J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [22] Lin H C, Lü Q, Wei H, et al. Quadrotor autonomous flight and three-dimensional dense reconstruction based on VI-SLAM[J]. Acta Optica Sinica, 2018, 38(7): 0715004.
林辉灿, 吕强, 卫恒, 等. 基于 VI-SLAM 的四旋翼自主飞行与三维稠密重构[J]. 光学学报, 2018, 38(7): 0715004.
- [23] Yang K, Fang W, Zhao Y, et al. Iteratively reweighted midpoint method for fast multiple view triangulation [J]. IEEE Robotics and Automation Letters, 2019, 4(2): 708-715.
- [24] Micusik B, Pajdla T. Structure from motion with wide circular field of view cameras[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(7): 1135-1149.
- [25] Ramalingam S, Lodha S K, Sturm P. A generic structure-from-motion framework[J]. Computer Vision and Image Understanding, 2006, 103(3): 218-228.
- [26] Schofield A, Ghasemi A, Vetterli M. Bound and conquer: improving triangulation by enforcing consistency [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(9): 2321-2326.
- [27] Lee S H, Civera J. Closed-form optimal two-view triangulation based on angular errors[C]//2019 IEEE International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea. New York: IEEE, 2019: 2681-2689
- [28] Enqvist O, Kahl F, Olsson C. Non-sequential structure from motion[C]//2011 IEEE International Conference on Computer Vision Workshops, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 264-271.